CHOICE WITH PROBABILISTIC REINFORCEMENT: EFFECTS OF DELAY AND CONDITIONED REINFORCERS

JAMES E. MAZUR

SOUTHERN CONNECTICUT STATE UNIVERSITY

Two experiments measured pigeons' choices between probabilistic reinforcers and certain but delayed reinforcers. In Experiment 1, a peck on a red key led to a 5-s delay and then a possible reinforcer (with a probability of .2). A peck on a green key led to a certain reinforcer after an adjusting delay. This delay was adjusted over trials so as to estimate an indifference point, or a duration at which the two alternatives were chosen about equally often. In all conditions, red houselights were present during the 5-s delay on reinforced trials with the probabilistic alternative, but the houselight colors on nonreinforced trials differed across conditions. Subjects showed a stronger preference for the probabilistic alternative when the houselights were a different color (white or blue) during the delay on nonreinforced trials than when they were red on both reinforced and nonreinforced trials. These results supported the hypothesis that the value or effectiveness of a probabilistic reinforcer is inversely related to the cumulative time per reinforcer spent in the presence of stimuli associated with the probabilistic alternative. Experiment 2 tested some quantitative versions of this hypothesis by varying the delay for the probabilistic alternative (either 0 s or 2 s) and the probability of reinforcement (from .1 to 1.0). The results were best described by an equation that took into account both the cumulative durations of stimuli associated with the probabilistic reinforcer and the variability in these durations from one reinforcer to the next.

Key words: reinforcer probability, reinforcer delay, conditioned reinforcement, adjusting schedule, key peck, pigeons

Two variables known to affect responding in choice situations are delay and probability of reinforcement. As the delay for one reinforcer increases, choice of that alternative becomes less likely, other things being equal (e.g., Green, Fisher, Perlow, & Sherman, 1981; Mazur, 1987; Rachlin & Green, 1972). Similarly, as the probability that a reinforcer will be delivered after a response decreases, choice of that alternative becomes less likely (e.g., Battalio, Kagel, & McDonald, 1985; Logan, 1965; Mazur, 1985).

One common method for examining the effects of these variables is to use a discrete-trials procedure, in which repeated choices between two alternatives are usually separated by some intertrial interval (ITI) between successive trials. Although delay and probability of reinforcement can be varied separately in this type of procedure, Rachlin, Logue, Gibbon, and Frankel (1986) proposed that probabilistic reinforcers function similarly to delayed reinforcers with reference to choice. Their theory, which was meant to apply to both human and nonhuman choice, stated that delivering a reinforcer on a probabilistic basis is equivalent to delivering a delayed reinforcer: If a reinforcer is not delivered after the subject's first response, the result is a delay between this first response and the eventual delivery of a reinforcer on a later trial. Rachlin et al. showed that, if the ITI is constant, the expected delay to reinforcement for any probabilistic reinforcer can be calculated with the following equation:

$$D = \frac{c + t}{p} - t, \qquad (1)$$

where D is the expected delay to a reinforcer delivered with a probability of p, t is the duration of the ITI, and c is time from the start of a trial until either a reinforcer is delivered or, on nonreinforced trials, until the trial ends and the ITI begins. Therefore, c includes the time it takes a subject to make the choice response plus any delay that may be imposed between the response and reinforcement (or between the response and trial termination on

nonreinforced trials). For example, suppose c is 2 s, t is 20 s, and p is .2. On average, it will take five trials to obtain a reinforcer, so Equation 1 states that D, the expected time to reinforcement, is 90 s (consisting of five trials of 2 s each, separated by four ITIs of 20 s each). The theory of Rachlin et al. therefore states that this probabilistic reinforcer will be equally preferred to a reinforcer delivered with certainty after a delay of 90 s.

Mazur (1989) concluded that although the fundamental idea behind the Rachlin et al. (1986) theory may be correct, Equation 1 did not correctly predict the choice behavior of pigeons. Mazur used a discrete-trials adjusting-delay procedure, in which subjects chose between a *standard alternative* and an *adjusting alternative*. For the standard alternative, the delay and probability of reinforcement were constant throughout a condition. For the adjusting alternative, a reinforcer was delivered on every trial, but the delay to reinforcement was systematically increased and decreased in 1-s increments several times a session, depending on the subject's choices. Adjusting the delay titrated an indifference point—a delay at which the two alternatives were chosen about equally often. In other words, by setting up a direct choice between a probabilistic reinforcer and a delayed reinforcer that was delivered with certainty, this procedure obtained estimates of D, the delay to reinforcement that was equivalent to a probabilistically delivered reinforcer. The obtained indifference points (which can be treated as estimates of D) were generally much shorter than predicted by Equation 1. Furthermore, large variations in the duration of the ITI had no effect on the obtained indifference points, whereas Equation 1 clearly predicts otherwise.

Mazur (1989) concluded that, at least for pigeons and this type of choice procedure, ITI duration had no measurable effect on choice. Removing ITI time from Equation 1 yields

$$D = \frac{c}{p}. \qquad (2)$$

The predictions of Equation 2 were closer to the actual results of Mazur's studies, but the predicted durations were still longer than the ones obtained. Mazur suggested that this might be because neither Equation 1 nor Equation 2 accommodates the variability in obtained delays with probabilistic reinforcers. For instance, with $p = .2$, the average number of trials between reinforcers will be five, but in individual instances the number will vary from one to many more than five trials between reinforcers. Because variable delays to reinforcement maintain more choice responding than a constant delay equal to the mean variable delay (Cicerone, 1976; Mazur, 1984), it is not surprising that the durations predicted by the above equations are too long. Both equations treat probabilistic reinforcers as if they imposed constant delays to reinforcement, whereas the actual delays are variable.

To incorporate the variable delays to reinforcement that occur with probabilistic reinforcers, a different equation is needed. Mazur (1984) found that the following hyperbolic equation made fairly accurate predictions for choices between reinforcers delivered after fixed and variable delays:

$$V = \sum_{i=1}^{n} P_i \left(\frac{1}{1 + KD_i}\right). \qquad (3)$$

V is the value of a reinforcement schedule composed of n different possible delays to reinforcement, where *value* refers to the schedule's ability to sustain choice responses. P_i is the probability that a delay of D_i seconds will occur on any given trial. (Note the distinction between P_i, the probability that a certain delay will occur, and p, the probability of reinforcement on a given trial.) K is a free parameter that determines how rapidly V declines with increasing values of D_i. In most cases, Mazur (1984) obtained good predictions if K, the free parameter, was simply set equal to 1.

Mazur (1989) used this same equation to predict indifference points for choices between certain and probabilistic reinforcers. The procedure entailed pigeons pecking either a green key to choose a certain reinforcer or a red key to choose a probabilistic reinforcer. After each peck on the green key, there was a delay (of adjustable duration) during which green houselights were lit, and then the reinforcer was delivered. After each peck on the red key, there was a 5-s delay during which red houselights were lit, and then a reinforcer might or might not be delivered. Each trial was followed by an ITI during which white houselights were lit.

Mazur (1989) found that Equation 3 more adequately predicted choice than either Equa-

tions 1 or 2, but only if D_i was defined as the cumulative time in the presence of the two stimuli defining the alternatives. That is, for the certain reinforcer, D_i was defined as the amount of time spent in the presence of the green response key and the green house-lights—the two stimuli that preceded this re-inforcer. For the probabilistic reinforcer, D_i was defined as the total amount of time spent in the presence of the red stimuli—the red key and the red houselights—from a subject's first choice of the probabilistic alternative until this alternative finally delivered a reinforcer (which often did not happen for several trials). For example, if the average response latency on the red key was 1 s, D_i would equal 6 s (1-s latency plus 5-s delay) if a reinforcer was delivered on the first trial that the probabilistic alternative was chosen. If a reinforcer was not delivered until the third trial with the probabilistic al-ternative, D_i would equal 18 s (three 1-s la-tencies plus three 5-s delays). To determine the overall value of V for the probabilistic al-ternative, Mazur measured such individual values of D_i across several sessions, then cal-culated a corresponding value of P_i for each value of D_i, and finally applied Equation 3 to estimate the overall value of the probabilistic alternative. To predict the indifference point, Equation 3 was then used to solve for an ad-justing delay that yielded the same value of V for the certain alternative.

Although this definition of D_i may seem counterintuitive, another study conducted by Mazur (1989) provided additional support for the definition. Some conditions were the same as those already described: For the probabi-listic alternative, the red houselights were lit for the 5-s delay that followed each peck on the red key, regardless of whether or not a reinforcer was delivered on that trial. In other conditions, however, the red houselights were omitted on nonreinforced trials, and the white houselights remained on during both the 5-s delay and the 55-s ITI that immediately fol-lowed. Although the probability of reinforce-ment was not changed, eliminating the red houselights on nonreinforced trials led to large and consistent decreases in the subjects' indif-ference points. For example, when the red houselights were present on both reinforced and nonreinforced trials and the probability of reinforcement was .2, the mean adjusting delay for the certain reinforcer was about 17 s. This

result means that, on average, the subjects were indifferent between a certain reinforcer deliv-ered after a 17-s delay and a reinforcer deliv-ered with a probability of .2 after a 5-s delay. However, when the red houselights were omit-ted on nonreinforced trials, the mean indiffer-ence point was only about 7 s. This decrease in the indifference point implied that V, the value of the standard alternative, had in-creased. The increase in V was predicted by Equation 3, because the omission of the red houselights on nonreinforced trials decreased the duration of D_i (as defined above). To put it simply, when subjects had to spend less time in the presence of the red houselights that were associated with the probabilistic alternative, they showed an increased preference for this alternative, even though the probability of re-inforcement and the actual time to reinforce-ment had not changed. Similar effects of stim-uli that precede a probabilistic reinforcer have been obtained by Dunn, Spetch, and their col-leagues in a series of experiments with con-current-chains procedures (Dunn & Spetch, 1990; Spetch, Belke, Barnet, Dunn, & Pierce, 1990).

In summary, this analysis suggests that the value of a probabilistic reinforcer depends on the cumulative amount of time spent in the presence of stimuli associated with that rein-forcer between a subject's initial choice re-sponse and the eventual delivery of the rein-forcer. These stimuli might be called conditioned reinforcers, because they preceded and were paired with the primary reinforcer, food. If this term is appropriate, then Equation 3 can be viewed as a method for predicting the strengths of these conditioned reinforcers. This approach resembles that of Vaughan (1985), who used a hyperbolic equation similar to Equation 3 to describe the relation between the durations of conditioned reinforcers and their values. The main difference is that Vaughan's equation did not take into account the variability in the durations of the condi-tioned reinforcers from one trial to the next.

The present experiments were designed to learn more about factors that can affect choices between certain and probabilistic reinforcers. In Experiment 1, the colors and durations of the stimuli that occurred after a choice of the probabilistic reinforcer were varied. The re-sults provided additional information about how the stimuli associated with a probabilistic

reinforcer can affect preference. Experiment 2 tested the generality of Equation 3 by determining whether it could be applied to situations in which the probabilistic alternative included no delay, or only a brief delay, between a choice response and reinforcement.

EXPERIMENT 1

The idea that the strength of a conditioned reinforcer is inversely related to its duration has many precedents (e.g., Dunn, Williams, & Royalty, 1987; Fantino, 1969, 1977; Killeen, 1982; Shull & Spear, 1987), although previous writers have not used the specific formulation presented in Equation 3. This notion may account for the increased preference for the probabilistic reinforcer when the red houselights were omitted on nonreinforced trials in Mazur's (1989) experiment. Another possibility, however, is that the colored lights were not conditioned reinforcers but simply discriminative stimuli that marked the delay between a choice response and the end of the trial. According to this hypothesis, preference for the probabilistic alternative increased when the red houselights were omitted on nonreinforced trials because stimuli associated with the ITI (the white houselights) occurred immediately after a response on the red key. Thus from the subject's perspective, a nonreinforced trial ended with a peck on the red key, even though from the experimenter's perspective there was still a 5-s delay after the key peck, but with white houselights rather than red houselights. According to this hypothesis, then, any stimuli that are uniquely associated with one alternative should be included as part of D_i, which can be viewed as a cumulative measure of trial duration. Any stimuli associated with the ITI are not included in D_i.

To evaluate the merits of these two hypotheses, Experiment 1 used a procedure similar to that of Mazur (1989), except that the houselight color on nonreinforced trials was red in some conditions, white in others, and blue in others, whereas the ITI always had a white houselight. The houselight was always red during the 5-s delay on reinforced trials for the probabilistic alternative. If the red keylight and houselights are conditioned reinforcers, they should be equally strong regardless of whether white houselights or blue houselights occur on nonreinforced trials. The calculation

of D_i should not include the durations of the blue houselights, because they were never paired with food and therefore cannot be considered conditioned reinforcers. According to this hypothesis, then, preference for the probabilistic alternative should be stronger with both the blue and white houselights than with the red houselights. If $K = 1$ and if response latencies average about 1 or 2 s, Equation 3 predicts that the indifference points should be almost twice as long in the conditions with the red houselights on nonreinforced trials. In contrast, if the discriminative-stimulus hypothesis is correct, the indifference points should be the same with the red and blue houselights, because both stimuli should clearly distinguish the trial from the ITI, and therefore both should contribute to D_i.

This experiment also included conditions in which the red houselights were presented for a full 60 s on nonreinforced trials (during both the 5-s delay and the 55-s ITI). If the red houselights are conditioned reinforcers, and if their value is inversely related to their duration, preference for the probabilistic alternative should decrease in these conditions because of the greatly increased duration of the red houselights. If $K = 1$ and if response latencies average about 1 or 2 s, Equation 3 predicts that indifference points should be about twice as long with the 60-s red houselights as with the 5-s red houselights. The discriminative-stimulus hypothesis makes no clear predictions for this situation. If the red houselights normally serve as discriminative stimuli that distinguish between the trial and the ITI, their roles as discriminative stimuli will be lost on nonreinforced trials when they are present during both the 5-s delay and the ITI.

METHOD

Subjects

Four White Carneau pigeons were maintained at about 80% of their free-feeding weights. All had previous experience with a variety of experimental procedures.

Apparatus

The experimental chamber was 30 cm long, 30 cm wide, and 33 cm high. Three response keys, each 2.5 cm in diameter, were mounted in the front wall of the chamber, 20.5 cm above the floor. A force of approximately 0.1 N was

required to operate each key, and each effective response produced a feedback click. A hopper below the center key provided controlled access to mixed grain, and when grain was available, the hopper was illuminated with two 6-W white lights. Eight 6-W lights (two white, two red, two green, and two blue) were mounted above the wire-mesh ceiling of the chamber. The chamber was enclosed in a sound-attenuating chamber containing a ventilation fan. A PDP-8® computer in another room was programmed in SuperSKED® to control the stimuli and record responses.

Procedure

The experiment consisted of 10 phases that included two or more replications of four different types of conditions. The four conditions differed only in the colors of the houselights that were presented on each nonreinforced trial with the standard alternative. The procedure in the 5-s blue condition (Phases 4 and 6) will be described first.

5-s blue condition (Phases 4 and 6). Each session lasted for 64 trials or for 80 min, whichever came first. Each block of four consecutive trials consisted of two forced trials followed by two choice trials. Figure 1 illustrates the possible sequences of events that could occur on a choice trial in these two conditions. Each trial was preceded by an ITI during which the white houselights were lit. The length of each ITI was set so that the total time between a choice response on one trial and the start of the next trial was 60 s. After the ITI, the center key was illuminated with white light to start a trial, and the white houselights remained on. A single peck on the center key, positioning the subject's head roughly equidistant from the two side keys, was required to begin the choice period. A peck on the center key darkened this key and illuminated the two side keys, one green and one red. Color positions (left or right key) varied randomly from trial to trial.

The red key was always the standard key, and the green key was the adjusting key. A peck on the red key extinguished both keylights. With a probability of .2, such pecks were followed by 5 s of red-houselight illumination, which was then extinguished during 2-s grain presentations, after which the white houselights were lit and the ITI began. Nonreinforced red-key choices were followed by
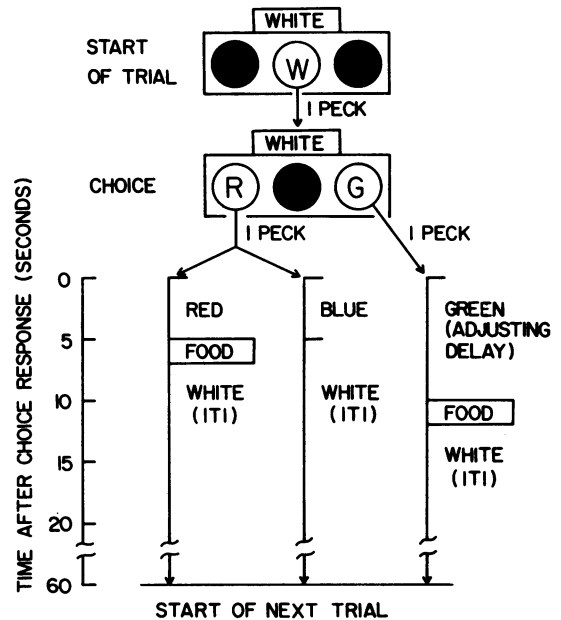

Fig. 1. A schematic diagram of the different sequences of events that could occur on a choice trial in the 5-s blue conditions of Experiment 1.

5-s blue houselights, and then the ITI began, during which the white houselights were lit.

If the green key was pecked during the choice period, both keylights were extinguished and the adjusting delay began, during which the green houselights were lit. The adjusting delay was always followed by 2-s access to grain (with all houselights off) and then the ITI (white houselights).

The procedure on forced trials was the same as on choice trials, except that only one side key was lit, red or green, and a peck on this key led to the appropriate delay. A peck on the opposite key, which was dark, had no scheduled effect. Of every two forced trials, one involved the red key and the other the green key. The temporal order of the red and green forced trials varied randomly.

After every four-trial block, the delay for the adjusting key was increased by 0.5 s if the adjusting key was chosen on both choice trials in the preceding block, decreased by 0.5 s if the standard key was chosen on both choice trials, and remained unchanged if each key was chosen once. In all three cases, this adjusting delay remained in effect for the next block of four trials. At the start of each session, the adjusting delay was determined by the above

Table 1

Order of phases and mean adjusting delays for each subject in Experiment 1. "Houselight colors" are the colors on nonreinforced red-key trials. The numbers of sessions required by each subject to meet the stability criteria are shown in parentheses. All durations are in seconds.

	Houselight colors		Mean adjusting delay (sessions)			
Phase	5-s delay	55-s ITI	Subject 1	Subject 2	Subject 3	Subject 4
1	red	white	18.24 (22)	14.55 (25)	13.24 (30)	16.66 (27)
2	white	white	7.66 (21)	8.40 (26)	7.43 (23)	9.33 (20)
3	red	white	16.40 (22)	12.80 (20)	9.50 (21)	16.06 (21)
4	blue	white	7.47 (27)	7.05 (24)	3.24 (24)	14.53 (21)
5	white	white	3.62 (24)	7.07 (26)	2.44 (22)	11.02 (20)
6	blue	white	6.51 (21)	6.48 (20)	9.30 (26)	11.03 (23)
7	red	white	24.45 (22)	12.70 (20)	5.92 (23)	20.42 (24)
8	red	red	20.79 (25)	22.10 (31)	10.69 (25)	20.15 (20)
9	red	white	16.00 (23)	12.09 (30)	7.10 (22)	16.69 (23)
10	red	red	15.67 (30)	22.16 (24)	14.56 (24)	12.72 (25)

rules as if it were a continuation of the preceding session.

These and all other phases lasted for a minimum of 20 sessions. After the minimum number of sessions, a phase was terminated for each subject individually when several stability criteria were met. To assess stability, each session was divided into two 32-trial blocks, and for each block the mean delay on the adjusting key was calculated. The results from the first two sessions of a phase were not used, and a phase was terminated when the following three criteria were met, using the data from all subsequent sessions: (a) Neither the highest nor the lowest single-block mean of a phase could occur in the last six blocks of the phase. (b) The mean adjusting delay across the last six blocks could not be the highest or the lowest six-block mean of the phase. (c) The mean delay of the last six blocks could not differ from the mean of the preceding six blocks by more than 10% or by more than 1 s (whichever was larger).

Other conditions. In the other three conditions, the procedure was the same following choices of the green key and following reinforced red-key choices. The only differences among conditions involved the colors of the houselights on nonreinforced red-key trials, and the names of the conditions refer to the stimuli on these trials. In the 5-s red condition (Phases 1, 3, 7, and 9), nonreinforced trials had the same houselight colors as reinforced trials (red during the 5-s delay and white during the subsequent ITI). The only difference between reinforced and nonreinforced red-key trials in this condition was that the 2-s reinforcer was

omitted and was replaced with an additional 2 s of ITI. These stimulus arrangements were repeated in four phases because they served as a baseline with which the effects of other stimuli could be assessed. In the 60-s white condition (Phases 2 and 5), no red houselights were present on nonreinforced red-key trials: The houselights were white during both the 5-s delay and the ITI (which therefore combined to form a continuous 60-s period with white houselights). Finally, in the 60-s red condition (Phases 8 and 10), the houselights on nonreinforced standard trials were red during both the 5-s delay and the ITI (which combined to form a continuous 60-s period with red houselights). The stimulus arrangements and number of sessions of exposure are summarized in Table 1.

RESULTS AND DISCUSSION

For each phase, all analyses were based on data from the six half-session blocks that satisfied the stability criteria. Averaged across subjects and phases, the mean response latency was 0.89 s on the standard key and 1.03 s on the adjusting key. The mean adjusting delays from the last six half-session blocks in each phase were used as estimates of the indifference points. Table 1 presents these means for each subject. A repeated-measures analysis of variance conducted with the data in Table 1 found a significant effect of condition, $F(9, 27) = 6.24$, $p < .001$. Figure 2 shows the means from the four different types of conditions, averaged across the two or four replications of each type. To provide a measure of within-subject variability in the adjusting delay, the
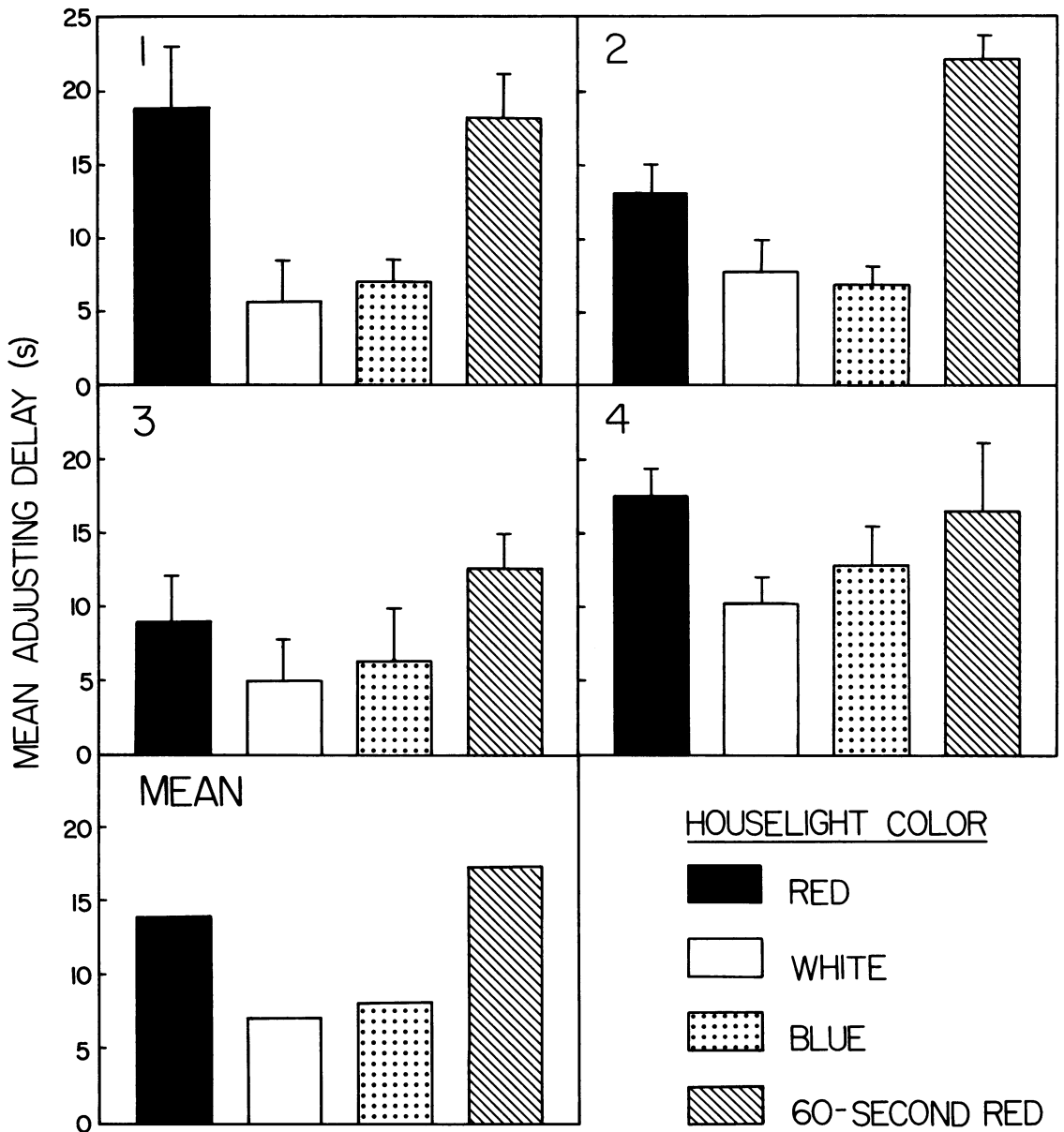
Fig. 2. For each subject and for the group mean, the mean adjusting delays are shown for the four different types of conditions in Experiment 1. Bars are standard deviations of the adjusting delays in each condition, calculated from the last six half-session blocks of each of the two (or four) replications of each condition. Houselight color refers to the color on nonreinforced red-key trials.

error bars in Figure 2 show the standard deviations of the adjusting delays in each condition, calculated from the last six half-session blocks of each of the two (or four) replications of each condition.

The overall pattern of results was quite similar for each of the 4 subjects. A comparison of the 5-s red and 60-s white conditions showed that Mazur's (1989) results were replicated—the indifference points were substantially shorter in the 60-s white condition. Averaged across subjects, the mean adjusting delay was 14.6 s in the 5-s red condition and 7.1 s in the 60-s white condition. Table 1 shows that these

differences were found for every subject in every comparison between a 5-s red phase and the 60-s white phase that preceded or followed it. A linear contrast found that the difference between these two types of conditions was highly significant, $F(1, 27) = 25.38, p < .001$.

Figure 2 also shows that the results from the 5-s blue condition, with an average adjusting delay of 8.2 s, were much closer to those of the 60-s white condition than the 5-s red condition. A linear contrast found a significant difference between the 5-s blue and 5-s red conditions, $F(1, 27) = 18.53, p < .001$, whereas another contrast found no significant difference between the 5-s blue and 60-s white conditions, $F(1, 27) = 3.62, p > .05$. These results indicate that the 5-s presentations of the blue houselights did not decrease preference for the probabilistic reinforcer in the same way the 5-s presentations of the red houselights did.

Compared to the other conditions, there was less consistency among subjects in the 60-s red condition. Figure 2 shows that the indifference points were longer in the 60-s red condition than in the 5-s red condition for only 2 of the 4 subjects. A linear contrast found no significant difference between the two types of conditions, $F(1, 27) = 0.40$.

These results provided partial support for the hypothesis that the red keylight and houselights served as conditioned reinforcers whose strengths can be derived from Equation 3. Comparisons of the 5-s red, 5-s blue, and 60-s white conditions strongly favored the conditioned-reinforcer hypothesis over the discriminative-stimulus hypothesis. Compared to the 60-s white condition, 5-s presentations of the red houselights on each nonreinforced trial decreased preference for the probabilistic alternative, whereas 5-s presentations of the blue houselights did not. This finding suggests that there is an important difference between a stimulus that simply distinguishes between trial and ITI (as did both the blue and red houselights) and one that is occasionally paired with the primary reinforcer (as were the red houselights but not the blue houselights). In terms of Equation 3, the results suggest that the blue houselights did not contribute to D_i, whereas the red houselights did. This is consistent with the hypothesis that only conditioned reinforcers (those that are occasionally followed by the primary reinforcer) contribute to D_i.

The results least supportive of the conditioned-reinforcer hypothesis came from the 60-s red condition. If the red houselights were conditioned reinforcers, substantially increasing their duration while providing no additional primary reinforcers should have decreased their reinforcing value. This prediction follows, not just from Equation 3, but from any theory that predicts an inverse relation between the duration of a conditioned stimulus and its value (e.g., Dunn et al., 1987; Shull & Spear, 1987). If the full 60-s durations of the red houselights on nonreinforced trials are included as part of D_i, Equation 3 predicts that the indifference points should have been about twice as long in the 60-s red condition as in the 5-s red condition. Although the overall comparison between these conditions was not significant, the results from 2 subjects supported this prediction. Table 1 shows that the indifference points for Subjects 2 and 3 were about twice as long in the two 60-s red phases as in the immediately preceding 5-s red phases. However, the results from the other 2 subjects showed no systematic differences between the 5-s red and 60-s red conditions.

One way the conditioned-reinforcer hypothesis could be modified to accommodate these results would be to propose that Subjects 1 and 4 discriminated the first 5 s of red houselights from the remainder. Note that the first 5 s of red houselights were indistinguishable on reinforced and nonreinforced trials. Continuations of the red houselights past 5 s were never followed by food. Hence it could be argued that, on nonreinforced trials, the first 5 s of red houselights were conditioned reinforcers (because of intermittent pairings with food) whereas the remaining portions of the 60-s interval were not (because food never followed). Although this reasoning might explain why indifference points were similar in the 5-s red and 60-s red conditions for these 2 subjects, it is clearly a post hoc attempt to explain results that were not anticipated, and there was no direct evidence that these 2 subjects formed temporal discriminations but the other 2 did not.

The conditioned-reinforcer hypothesis does not provide the only possible explanation of the present results. The results are also consistent with the view that information about whether or not a reinforcer will be delivered can itself serve as a reinforcer (e.g., Bower, McLean, & Meacham, 1966; Hendry, 1969).

In the present experiment, a choice of the red key provided immediate information about the presence or absence of a reinforcer in the 5-s blue and 60-s white conditions but not in the 5-s red or 60-s red conditions.

Except for the questions raised by the between-subject differences in the 60-s red condition, the predictions of Equation 3 have fared reasonably well. This equation has made predictions for five previous experiments on probabilistic reinforcers (Mazur, 1985, 1989) that were substantially more accurate than those of alternative approaches (Equations 1 and 2). Assuming that only conditioned reinforcers contribute to D_i, Equation 3 can also account for the differences among the 5-s red, 5-s blue, and 60-s white conditions of the present experiment. As Equation 3 predicted, indifference points were, on average, almost twice as long in the 5-s red condition as in the other two conditions. Experiment 2 was designed to test some additional predictions of Equation 3.

EXPERIMENT 2

According to the theory summarized by Equation 3, preference for a probabilistic reinforcer is inversely related to the time spent in the presence of stimuli associated with that reinforcer. In all of the previous studies that were used to test this theory, the probabilistic alternative included a 5-s delay between a choice response and the outcome of the trial (reinforcement or nonreinforcement). This delay meant that D_i would always be greater than 5 s (greater because some time was spent in the presence of the keylight as well), especially when there were several trials with the probabilistic alternative before a reinforcer was delivered. However, suppose the probabilistic alternative included no delay between a choice response and either reinforcement or trial termination. For this situation, Equation 3 makes a prediction that may seem counterintuitive: Because D_i would include only the time spent in the presence of the keylight before a choice response was made, Equation 3 predicts that the equivalent adjusting delay for a certain reinforcer should increase only slightly as p_s, the probability of reinforcement for the standard alternative, decreases. For example, in one condition of Experiment 2, p_s was .1, there was no delay for the probabilistic alternative,

and the ITI was 30 s. If $K = 1$, and if response latencies average 1 s, Equation 3 predicts that the adjusting delay at the indifference point should be 3.8 s. This predicted duration may seem surprisingly short, considering that an average of 10 trials with the probabilistic alternative, each separated by at least 30 s, would be required for each delivery of the probabilistic reinforcer.

To test the predictions of Equation 3, this experiment included four conditions, each with a different value of p_s, that had no delay between a choice response and either reinforcement or trial termination. Four other conditions included a brief (2 s) delay.

METHOD

Subjects and Apparatus

Four experimentally naive White Carneau pigeons were maintained at about 80% of their free-feeding weights.

The experimental chamber was identical to that of Experiment 1, except that all the lights in the chamber used 2-W bulbs rather than 6-W bulbs. All stimuli were controlled and responses recorded by an IBM-compatible personal computer using the Medstate® programming language.

Procedure

Subjects were first trained to eat from the food hopper; then an autoshaping procedure was used to establish pecking at the center key. Next, another autoshaping procedure was used to establish pecking at all three keys. Once each subject learned to peck promptly at any key that was lit, Condition 1 of the experiment began.

The experiment used the same type of adjusting-delay procedure as in Experiment 1, except for the following changes. The key and houselight colors were the opposite of those in Experiment 1: The green keylight and houselights were associated with the standard alternative (with a probabilistic reinforcer), and the red keylight and houselights were associated with the adjusting alternative (with a certain reinforcer). The positions of the two key colors did not vary randomly: The green key was always on the left, and the red key was always on the right. The duration of each reinforcer was 3 s, the total time between a choice response and the start of the next trial was 30

Table 2

Order of conditions in Experiment 2 and the number of sessions required by each subject to meet the stability criteria.

Condi-tion	Stan-dard delay (s)	Stan-dard proba-bility	Number of sessions			
			Sub-ject 1	Sub-ject 2	Sub-ject 3	Sub-ject 4
1	6	1.0	25	27	26	26
2	0	.2	30	25	26	26
3	0	.5	12	12	13	15
4	0	.1	14	15	13	16
5	2	.1	13	12	15	14
6	2	.5	13	15	13	12
7	2	.2	15	15	12	23
8	2	1.0	15	13	12	14
9	0	1.0	12	13	13	12

s, the maximum possible adjusting delay was 26 s, and sessions ended after 64 trials or 50 min, whichever came first. Changes in the adjusting delay were made in 1-s increments rather than 0.5-s increments. Other details of the procedure, such as the stability criteria and the rules for changing the adjusting delay, were the same as in Experiment 1.

The experiment included nine conditions. The first condition, which continued for a minimum of 25 sessions and until the stability criteria were satisfied, was used to train the subjects with the adjusting procedure and to measure possible key color or position bias. For both the standard and adjusting alternatives, each trial ended with reinforcement. After a response on the green key (the standard alternative), there was a 6-s delay during which the green houselights were lit, followed by reinforcement. After a response on the red key (the adjusting alternative), the red houselights were lit during the adjusting delay, which was followed by reinforcement.

Condition 2 lasted for a minimum of 25 sessions, and the other seven conditions lasted for a minimum of 12 sessions each, until the stability criteria were satisfied. Table 2 shows the number of sessions needed to satisfy these criteria, as well as the standard delay and probability of reinforcement for each condition. In four conditions, the standard delay was 0 s, and the standard probability of reinforcement was varied between .1 and 1.0. In the other four conditions, the standard delay was 2 s, and the standard probability of reinforcement was varied between .1 and 1.0. In the

conditions with the 2-s delay, each choice response on the green (standard) key was followed by a 2-s illumination of the green houselights, regardless of whether or not a reinforcer was to be delivered. In the conditions with the 0-s delay, each choice response on the green key was followed by either a reinforcer or simply the darkening of the response keys and continuation of the white houselights.

RESULTS AND DISCUSSION

All analyses were based on the six half-session blocks that satisfied the stability criteria in each condition. Averaged across subjects and conditions, the mean response latency on the standard key was 1.04 s and was 1.10 s on the adjusting key. Figure 3 shows the mean adjusting delays ($\pm SEM$) for each subject in each condition. This figure shows the effects of both standard delay and standard probability. In every case, for any given probability of reinforcement, the mean adjusting delay was greater with a 2-s standard delay than with the 0-s standard delay. For a given standard delay, the adjusting delay decreased steadily as the standard probability of reinforcement increased (with one reversal for Subject 3). Figure 3 also shows the results from Condition 1, which had a standard delay of 6 s. The equivalent adjusting delays were slightly greater than 6 s for each of the 4 subjects, and averaged 7.2 s. These results therefore indicated a slight bias toward the adjusting (or red) key. At least two other studies with the adjusting-delay procedure (Mazur, 1984; Mazur, Snyderman, & Coe, 1985) have found a bias toward the adjusting key (which was green rather than red in those studies).

Although there was some variability among subjects in the durations of the adjusting delays, for all subjects these durations were in the general range predicted by Equation 3. In the conditions with standard delays of 0 s, the equivalent adjusting delays never exceeded 10 s, even when the probability of reinforcement was .1. To provide a more precise test of Equation 3, its predictions were calculated for the different conditions of this experiment and compared to the group means. The calculations took into account the mean response latencies on the standard and adjusting keys. They also took into account the bias toward the adjusting key observed in Condition 1. Because the mean adjusting delay of 7.2 s in
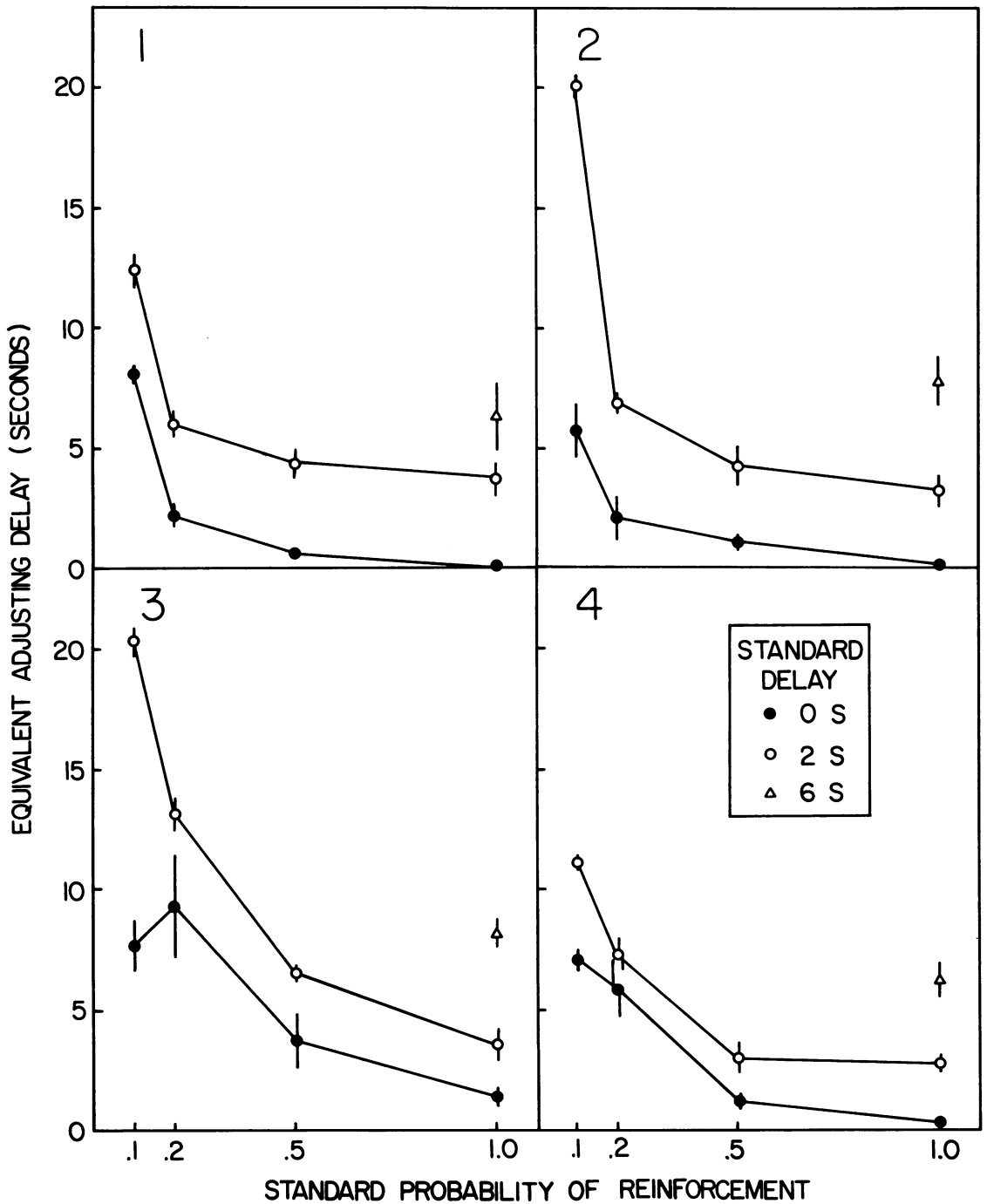
Fig. 3. The mean adjusting delays are shown for each subject from the nine conditions of Experiment 2, plotted as a function of the probability of reinforcement on the standard key. The error bars show ±1 *SEM* for all cases in which the standard errors extend beyond the size of the symbol. The standard errors were calculated from the six half-session blocks that satisfied the stability criteria in each condition.
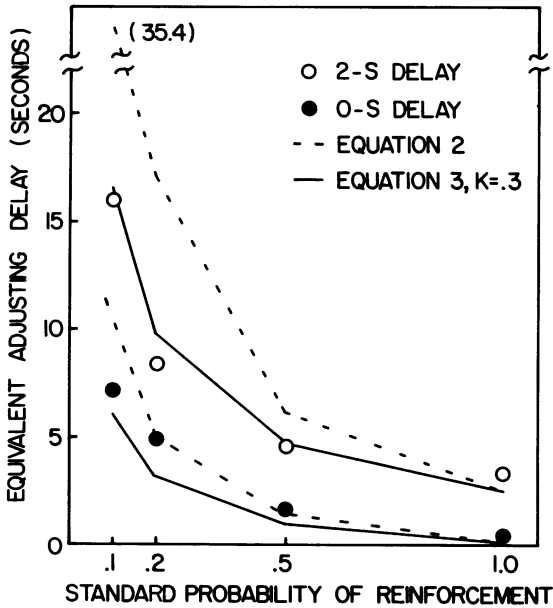
Fig. 4. The group means (points) from Experiment 2 are compared to the predictions of Equation 2 (broken lines) and Equation 3 (solid lines). For Equation 3, the predictions are based on the best fitting value of K. For each equation, the upper line shows predictions for conditions with 2-s standard delays, and the lower line shows predictions for conditions with 0-s standard delays.

Condition 1 was 20% longer than the 6-s standard delay, all predictions of Equation 3 were increased by 20%. This 20% increase was used (instead of simply adding 1.2 s to each prediction, for example) because Mazur's (1984) experiment with the adjusting-delay procedure found that bias was proportional to the duration of the adjusting delay. In calculating the predictions of Equation 3, K was treated as a free parameter, and a value of K was chosen to maximize the correspondence between predictions and the mean indifference points for the group, using a least-squares criterion.

Figure 4 compares the group means from each condition to the predictions of Equation 3 with $K = 0.3$, the best fitting value of K. There was a good correspondence between the predictions and the group means, and Equation 3 accounted for 95.5% of the variance in the group means. In addition, the overall pattern of results was close to what the equation predicted—a curvilinear increase in the indifference points as p_s decreased, both with the 0-s and 2-s standard delays. For comparison, Figure 4 also shows the predictions of Equa-

tion 2, adjusted to take into account the mean response latencies and the bias toward the adjusting key. The predictions of Equation 2 were also fairly accurate with high reinforcement probabilities, but the predicted durations were much too long with the lower probabilities and a standard delay of 2 s. Because Equation 2 includes no free parameters, it cannot be expected to be as accurate as Equation 3, which has K as a free parameter. However, the greater accuracy of Equation 3 cannot be attributed entirely to the presence of the free parameter, because with all values of K greater than 0, this equation predicts shorter indifference points than Equation 2. (With $K = 0$, the predictions of the two equations become identical. Of course, setting $K = 0$ means that delay is assumed to have no effect on value.) Thus the main problem for Equation 2 was the same as in previous studies (Mazur, 1989): As the probability of reinforcement decreased, Equation 2 predicted indifference points that were longer than those actually observed. The predictions of Equation 1 are not shown in Figure 4, but because this equation includes the durations of the ITIs, it predicts much longer indifference points. In fact, except for the conditions with $p_s = 1$, Equation 1 predicts that the adjusting delays should have reached their maximum possible duration of 26 s. As in previous studies, the predictions of Equation 3 were the most accurate.

The predictions of Equations 2 and 3 were also compared to the results from the individual subjects, in the same way that they were compared to the group results. With Equation 3, the best fitting values of K were 0.3, 0.6, 0.2, and 0.6, for the 4 subjects, respectively. Equation 3 accounted for 80%, 92%, 87%, and 75% of the variance in the data points for the 4 subjects, respectively. Except for the somewhat lower percentages, the overall patterns of predictions and results for the individual subjects were similar to that shown for the group means in Figure 4. These values of K were lower than the mean value of 1.0 obtained by Mazur (1984), but the best fitting values for individual subjects in that study were 0.4, 0.7, 1.0, and 2.6. Given this variability, the estimates of K for individual subjects were not reliably different in the two studies.

The predictions of Equation 2 for individual subjects were too long for the conditions with low reinforcement probabilities and a standard

delay of 2 s, just as they were for the group results. Because of the inaccurate predictions in these conditions, the percentages of variance accounted for were less than 0% for all 4 subjects (meaning that the variance from the predictions of Equation 2 was greater than the simple variance from the mean of the nine indifference points). In summary, for both individual subjects and the group means, the predictions of Equation 3 were the most accurate.

GENERAL DISCUSSION

These two experiments supported the hypothesis that the value of a probabilistic reinforcer is determined by the total time spent in the presence of stimuli associated with that reinforcer before the reinforcer is actually delivered. Inasmuch as duration was shown to be an important variable, the results are consistent with the general view of Rachlin et al. (1986) that probabilistic reinforcers are functionally equivalent to delayed reinforcers. However, as in previous studies with pigeons (Mazur, 1989), the present experiments showed that Equation 1 is not suitable, at least with this type of procedure, because ITI time does not appear to affect the value of a probabilistic reinforcer. Only time spent in the presence of stimuli uniquely associated with the probabilistic reinforcer (the red lights in Experiment 1 and the green lights in Experiment 2) influenced the preference for that reinforcer, as measured by the equivalent adjusting delays.

If the value of a probabilistic reinforcer is determined by the cumulative duration of reinforcer-related stimuli, then decreasing the probability of reinforcement should have little effect when the durations of these stimuli are short on every trial. Experiment 2 tested this prediction by removing the delay between a choice response and either reinforcement or trial termination. Thus, the only reinforcer-related stimulus was the green keylight, which was present only for the brief time (averaging just over 1 s) between the onset of the keylights and the subject's response. Consistent with the prediction, indifference points changed by only a few seconds as the probability of reinforcement varied from 1.0 to .1, so that with a probability of .1, the average indifference point was only 7.2 s.

Experiment 2 also provided support for the view that the variability in the durations of these stimuli from one reinforcer to the next must be taken into account. The predictions of Equation 2, which considers only average durations, were consistently too long for conditions in which the probability of reinforcement was low (.2 or less). Equation 3, which takes into account the variability in D_i from one reinforcer to the next, accounted well for the pattern of results in this experiment. Equation 3 has also provided more accurate predictions than Equation 2 for several previous studies on probabilistic reinforcement (Mazur, 1985, 1989). It may therefore offer a useful method for predicting the value of probabilistic reinforcers as well as those delivered after fixed or variable delays.

Experiment 1 tested the hypothesis that D_i is best interpreted as a measure of the duration of conditioned reinforcers that are associated with the probabilistic reinforcer. An alternative hypothesis was that D_i should be interpreted as a measure of any discriminative stimuli uniquely associated with the probabilistic reinforcer, whether or not they can be called conditioned reinforcers. The results from the 60-s red condition were ambiguous, but the results from the 5-s blue condition clearly supported the conditioned-reinforcer hypothesis. The 5-s blue houselights were designed to be discriminative stimuli but not conditioned reinforcers: They were uniquely associated with the probabilistic alternative but cannot be considered conditioned reinforcers because they were never paired with the primary reinforcer, food. The results indicated that the durations of these blue houselights should not be included as part of D_i, because the indifference points in this condition were significantly shorter than those in the condition in which 5-s red houselights occurred on nonreinforced trials. In fact, the indifference points from the 5-s blue condition were similar to those from the 60-s white condition. This suggests that the blue houselights did not affect the value of the probabilistic alternative any more than did the white houselights of the ITI. This finding supports the conditioned-reinforcer hypothesis because it suggests that only stimuli that are occasionally paired with the primary reinforcer contribute to D_i and thereby affect the value of the probabilistic reinforcer.

The results from the 60-s red condition were

ambiguous because two subjects behaved as the conditioned-reinforcer hypothesis predicted, but the other two did not. Some possible explanations of these results have already been discussed, but no firm conclusions can be drawn without additional research on the effects of such prolonged stimuli.

The idea that the strength of a conditioned reinforcer is inversely related to its duration is not new. According to what Fantino (1977) called the *reinforcement-density hypothesis,* the strength of a conditioned reinforcer is "determined by the rate of primary reinforcement in its presence" (p. 314). Vaughan's (1985) theory applied the reinforcement-density hypothesis in its calculation of the value of a conditioned reinforcer. Duration is also an important variable in Fantino's (1969) delay-reduction theory, which states that the strength of a conditioned reinforcer depends on the reduction in time to primary reinforcement signaled by the onset of the conditioned reinforcer. These theories consider only the *average time* to primary reinforcement or its inverse, the *average rate* of primary reinforcement. However, Shull and Spear (1987) proposed an hypothesis about the strength of conditioned reinforcers that is closer to the view presented here. They suggested using a formula similar to Equation 3 to calculate the conditioned reinforcing strength of a stimulus that is followed by a primary reinforcer after variable delays.

The theoretical approach described in this paper differs from previous theories of conditioned reinforcement in one noteworthy way. Although the duration of a conditioned reinforcer is used in Equation 3, this duration is measured cumulatively—from the first trial on which the probabilistic alternative is chosen until the primary reinforcer is delivered, even though other stimuli (those associated with the ITI or with an alternative source of reinforcement) may occur in between. The measurement of cumulative durations may seem counterintuitive, but the best argument for using this approach is that it seems to work. This application of Equation 3 has provided a good account of the present results and previous results (Mazur, 1985, 1989). Other plausible models of probabilistic reinforcement (those represented by Equations 1 and 2) were considerably less accurate. However, many other theories relating probability, delay, and conditioned reinforcement could be formulated,

and whether measuring cumulative stimulus durations is the most satisfactory approach to the problem of probabilistic reinforcement remains to be seen.

REFERENCES

Battalio, R. C., Kagel, J. H., & McDonald, D. N. (1985). Animals' choices over uncertain outcomes: Some initial experimental results. *American Economic Review, 75,* 597–613.

Bower, G., McLean, J., & Meacham, J. (1966). Value of knowing when reinforcement is due. *Journal of Comparative and Physiological Psychology, 62,* 184–192.

Cicerone, R. A. (1976). Preference for mixed *versus* constant delay of reinforcement. *Journal of the Experimental Analysis of Behavior, 25,* 257–261.

Dunn, R., & Spetch, M. L. (1990). Choice with uncertain outcomes: Conditioned reinforcement effects. *Journal of the Experimental Analysis of Behavior, 53,* 201–218.

Dunn, R., Williams, B., & Royalty, P. (1987). Devaluation of stimuli contingent on choice: Evidence for conditioned reinforcement. *Journal of the Experimental Analysis of Behavior, 48,* 117–131.

Fantino, E. (1969). Choice and rate of reinforcement. *Journal of the Experimental Analysis of Behavior, 12,* 723–730.

Fantino, E. (1977). Conditioned reinforcement: Choice and information. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 313–339). Englewood Cliffs, NJ: Prentice-Hall.

Green, L., Fisher, E. B., Jr., Perlow, S., & Sherman, L. (1981). Preference reversal and self control: Choice as a function of reward amount and delay. *Behaviour Analysis Letters, 1,* 43–51.

Hendry, D. P. (1969). Reinforcing value of information: Fixed-ratio schedules. In D. P. Hendry (Ed.), *Conditioned reinforcement* (pp. 300–341). Homewood, IL: Dorsey Press.

Killeen, P. R. (1982). Incentive theory: II. Models for choice. *Journal of the Experimental Analysis of Behavior, 38,* 217–232.

Logan, F. A. (1965). Decision making by rats: Uncertain outcome choices. *Journal of Comparative and Physiological Psychology, 59,* 246–251.

Mazur, J. E. (1984). Tests of an equivalence rule for fixed and variable reinforcer delays. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 426–436.

Mazur, J. E. (1985). Probability and delay of reinforcement as factors in discrete-trial choice. *Journal of the Experimental Analysis of Behavior, 43,* 341–351.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Hillsdale, NJ: Erlbaum.

Mazur, J. E. (1989). Theories of probabilistic reinforcement. *Journal of the Experimental Analysis of Behavior, 51,* 87–99.

Mazur, J. E., Snyderman, M., & Coe, D. (1985). Influences of delay and rate of reinforcement on discrete-

trial choice. *Journal of Experimental Psychology: Animal Behavior Processes, 11,* 565–575.

Rachlin, H., & Green, L. (1972). Commitment, choice, and self-control. *Journal of the Experimental Analysis of Behavior, 17,* 15–22.

Rachlin, H., Logue, A. W., Gibbon, J., & Frankel, M. (1986). Cognition and behavior in studies of choice. *Psychological Review, 93,* 33–45.

Shull, R. L., & Spear, D. J. (1987). Detention time after reinforcement: Effects due to delay of reinforcement? In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on*

reinforcement value (pp. 187–204). Hillsdale, NJ: Erlbaum.

Spetch, M. L., Belke, T. W., Barnet, R. C., Dunn, R., & Pierce, W. D. (1990). Suboptimal choice in a percentage-reinforcement procedure: Effects of signal condition and terminal-link length. *Journal of the Experimental Analysis of Behavior, 53,* 219–234.

Vaughan, W., Jr. (1985). Choice: A local analysis. *Journal of the Experimental Analysis of Behavior, 43,* 383–405.